

Cross-Country Analyses of Economic Growth: An Econometric Survey *

Fernanda Llussá [†]

Faculdade de Ciência e Tecnologia, Universidade Nova de Lisboa

and

INOVA Research Center.

September 2007.

Abstract

This paper reviews the econometric methodology on panel data estimation and testing as applied to the study of convergence in growth empirics. The concept of absolute convergence states that the poorer economies should be growing at a faster rate, catching up the richer ones. The empirical failure of absolute convergence resulted in the development of alternative theories to explain long-term growth: the endogenous growth theories and the conditional convergence, the idea that countries may have different steady-states and it is the distance from their own steady-state that determines the rate of economic growth. This paper focuses on conditional convergence and its empirical testing. It discusses and compares the different econometric methodologies used in cross-section and panel data studies of conditional convergence. Also presented are the empirical results obtained by the various authors.

Key Words: *General Method of Moments, Conditional Convergence, Panel Data.*

JEL Classification: C3, C31, C33..

*I would like to thank Ekaterini Kyriazidou, Luisa Lambertini, Kenneth Sokoloff and Carlos for comments and support for this work. All errors remain my own.

[†]Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, 2829-516 Caparica, Portugal. e-mail: fajl@fct.unl.pt

1 Introduction

Convergence is a simple but powerful idea: do poor countries tend to grow at faster rates than rich ones? In a progressively more interconnected world, this is an important empirical question with broad implications. As will be evident from this paper, the answer to the question is no, in absolute terms, and maybe, conditional on other country characteristics.

The basic theoretical framework used to approach this question has been the Solow (1956) model, further developed by Cass (1965) and Koopmans (1965). The main force driving economic growth in this model is capital accumulation. When, as is likely in poorer countries, the capital-labor ratio is low, the marginal product of capital is high and investment leads to large increases in output. As the economy gets richer and its capital-labor ratio increases, the growth rate of output tends to decrease.

The concept of absolute convergence states that the poorer economies should be growing at a faster rate, catching up the richer ones. In the background is the idea that economies are all converging to the same long-run steady-state so that economies more distant from the steady-state tend to grow faster. Empirical work found no evidence of absolute convergence for large cross-sections of countries and the concept was only verified for smaller and more homogeneous subgroups, such as the OECD. The empirical failure of absolute convergence resulted in the development of alternative theories to explain long-term growth. On the one hand, endogenous growth theories relied on the fact that capital accumulation per se was not enough to explain growth. Instead, spillovers from capital accumulation facilitate higher rates of technological progress, capable of explaining why richer countries may grow faster than poorer ones. A second response to the empirical failure of absolute convergence was conditional convergence, the idea that countries may have different steady-states and it is the distance from their own steady-state that determines the rate of economic growth. This paper focuses on conditional convergence and its empirical testing. Conditional convergence is thus able to reconcile the empirical facts with a modified, simple Solow model, in which countries converge to different steady-states.

This paper is organized in the following way: the second section presents a literature review of each of the three theoretical models used to study convergence, namely the Solow model, the Augmented Solow model and the modified Solow model à la Barro. The third section discusses and compares the different econometric methodologies used in cross-section and panel data studies of conditional convergence. Also presented are the empirical results obtained by the various authors. The fourth section concludes with suggestions for further research.

2 Theoretical Framework

There are three basic models that have been proposed to explain differences in growth rates across countries: the Solow model, the Augmented Solow model and the modified Solow model à la Barro.

2.1 The Solow Model

The key assumption of the Solow model is that there are diminishing marginal returns to capital. Countries with higher capital stock will grow at a progressively lower rate until they reach their steady-state level of income. At the steady-state output per capita, capital stock and consumption grow at the same constant rate, equal to the exogenous growth rate of technological progress. The model is based on the assumption of a Cobb-Douglas production function with labor augmenting technological progress:¹

$$Y(t) = K(t)^\alpha [A(t)L(t)]^{1-\alpha}, 0 < \alpha < 1 \quad (1)$$

where t is time, $Y(t)$ output, $K(t)$ capital stock, $L(t)$ labor, $A(t)$ technological progress and α the share of capital on output.²

Assume that the population growth rate is given by n and the rate of technological progress by g . Thus,

$$L(t) = L(0)e^{nt} \quad (2)$$

$$A(t) = A(0)e^{gt} \quad (3)$$

The main equation of the Solow model used for estimation purposes is equation (4) below.³

$$\begin{aligned} \ln \hat{y}(t_2) - \ln \hat{y}(t_1) &= (1 - e^{-\lambda\tau}) \frac{\alpha}{(1 - \alpha)} \ln s \\ &- (1 - e^{-\lambda\tau}) \frac{\alpha}{(1 - \alpha)} \left[\ln(n + g + \delta) - (1 - e^{-\lambda\tau}) \ln \hat{y}(t_1) \right] \end{aligned} \quad (4)$$

where $s = \frac{\dot{K}(t) + \delta K(t)}{Y(t)}$, is the savings rate, $\lambda = (n + g + \delta)(1 - \alpha)$ is the rate of convergence and δ is the depreciation rate.

¹See the Appendix for a detailed mathematical derivation of all equations below.

²In a Cobb-Douglas the share of capital on output is equal to the elasticity of output in relation to capital given by

$$d \ln Y(t) / d \ln K(t)$$

³Equation (71) in the Appendix.

Note that in equation (4), the steady-state values of \hat{k}^* and \hat{y}^* are ⁴:

$$\hat{k}^* = \left(\frac{s}{n + g + \delta} \right)^{\frac{1}{1-\alpha}} \quad (5)$$

$$\hat{y}^* = \left(\frac{s}{n + g + \delta} \right)^{\frac{\alpha}{1-\alpha}} \quad (6)$$

We can get the speed of convergence by linearizing around the steady-state \hat{y}^* :

$$\dot{\hat{y}}(t) \simeq \left(\frac{\partial \hat{y}(t)}{\partial \hat{y}(t)} \Big|_{\hat{y}(t)=\hat{y}^*} \right) (\hat{y}(t) - \hat{y}^*) \quad (7)$$

Note that, in equation (4)

$$\hat{y}(t) = \frac{Y(t)}{A(t)L(t)} \quad (8)$$

i.e $\hat{y}(t)$ is output per “effective-labor ” (the total factor productivity $A(t)$ times total labor $L(t)$ in the economy).

In order to proceed it is necessary to rewrite equation (4) in terms of per capita output as the variable $A(t)$ is unobserved. We know that:⁵

$$\ln \hat{y}(t) = \ln y(t) - \ln A(t) \quad (9)$$

As far as the technological progress term $A(t)$ goes, there is a key difference between the assumptions of Mankiw, Romer and Weil (1992) - from now on referred as MRW (1992), and those of Caselli, Esquivel and Lefort (1996) - from now on referred as CEL (1996) - and Islam (1995).

(i) MRW (1992) estimate a cross-section and assume that $\ln A(t) = \ln A(0)$ which does not vary with time so that

$$\ln A(0) = a + \epsilon \quad (10)$$

where a is a constant and ϵ a country-specific shift (shock term). From equations (4), (9) and (10) MRW (1992) obtain (11) which they use for estimation purposes.

$$\begin{aligned} \ln y(t_2) - \ln y(t_1) &= (1 - e^{-\lambda\tau}) \frac{\alpha}{(1 - \alpha)} \ln s \\ &- (1 - e^{-\lambda\tau}) \frac{\alpha}{(1 - \alpha)} \ln(n + g + \delta) \\ &+ (1 - e^{-\lambda\tau}) a - (1 - e^{-\lambda\tau}) \ln y(t_1) + (1 - e^{-\lambda\tau}) \epsilon \end{aligned} \quad (11)$$

⁴In the steady-state, $\dot{\hat{k}}(t) = 0$. The Appendix presents how to obtain the steady-state values of \hat{k}^* and \hat{y}^* .

⁵Where $A(t) = A(0)e^{gt}$.

Note that the intercept $(1 - e^{-\lambda\tau})a$ is constant in the cross-section. MRW (1992) are assuming that countries have the same level of technological progress $A(0)$ and recognize the implications of this not being true:

“If countries have permanent differences in their production functions—that is, different $A(0)$ ’s - then, these $A(0)$ ’s would be positively correlated with initial income. Hence, variations in $A(0)$ would bias the coefficient on initial income toward zero (and would potentially influence the other coefficients as well). In other words, permanent cross-country differences in the production function would lead to differences in initial incomes uncorrelated with subsequent growth rates and, therefore, would bias the results against finding convergence” (MRW (1992), p. 424).

(ii) Islam (1995) and CEL (1996) disagree with the assumption of a common production function across countries. They assume instead that the level as well as the rate of technological progress differ across countries. Thus, in place of equation (10), the authors assume that:

$$\ln A(t) = \ln A(0) + gt \quad (12)$$

From equations (4), (9) and (12) the authors obtain (13) which is used for estimation purposes:

$$\begin{aligned} \ln y(t_2) = & (1 - e^{-\lambda\tau}) \frac{\alpha}{(1 - \alpha)} \ln s - (1 - e^{-\lambda\tau}) \frac{\alpha}{(1 - \alpha)} \ln(n + g + \delta) \\ & + e^{-\lambda\tau} \ln y(t_1) + (1 - e^{-\lambda\tau}) \ln A(0) + g(t_2 - e^{-\lambda\tau}t_1) \end{aligned} \quad (13)$$

Rewriting (13) in panel data notation we get:

$$y_{it} = \gamma y_{i,t-1} + \sum_{j=1}^2 \beta_j x_{it}^j + \eta_t + \mu_i + v_{it} \quad (14)$$

where

$$\begin{aligned} y_{it} &= \ln y(t_2), \quad y_{i,t-1} = \ln y(t_1), \quad \gamma = e^{-\lambda\tau} \\ \beta_1 &= (1 - e^{-\lambda\tau}) \frac{\alpha}{(1 - \alpha)}, \quad \beta_2 = -(1 - e^{-\lambda\tau}) \frac{\alpha}{(1 - \alpha)} \\ x_{it}^1 &= \ln s, \quad x_{it}^2 = \ln(n + g + \delta) \\ \eta_t &= g(t_2 - e^{-\lambda\tau}t_1), \quad \mu_i = (1 - e^{-\lambda\tau}) \ln A(0), \quad v_{it} = error \end{aligned}$$

with $E(v_{it}) = 0$. Note that μ_i is the time-invariant individual country effect and v_{it} is the error term that varies with countries and time periods.

2.2 The Augmented Solow Model

The Augmented Solow model includes human capital in the production function:

$$Y(t) = K(t)^\alpha H(t)^\beta [A(t)L(t)]^{1-\alpha-\beta}, 0 < \alpha < 1 \quad (15)$$

where $H(t)$ is the stock of human capital.

The main equation of the Augmented Solow model used for estimation purposes is equation (16) below.

$$\begin{aligned} \ln \hat{y}(t_2) - \ln \hat{y}(t_1) &= (1 - e^{-\lambda\tau}) \frac{\alpha}{(1 - \alpha - \beta)} \ln s_k \\ &+ (1 - e^{-\lambda\tau}) \frac{\beta}{(1 - \alpha - \beta)} \ln s_h - (1 - e^{-\lambda\tau}) \frac{\alpha + \beta}{(1 - \alpha - \beta)} \ln(n + g + \delta) \\ &- (1 - e^{-\lambda\tau}) \ln \hat{y}(t_1) \end{aligned} \quad (16)$$

where s_k is the savings in physical capital, s_h the savings in human capital, λ is the rate of convergence, δ is the rate of depreciation and $\tau = t_2 - t_1$.

Again, before proceeding to estimate the model, it is necessary to rewrite equation (16) in terms of per capita output. Note again that for MRW (1992):

$$\ln A(0) = a + \epsilon \quad (17)$$

On the other hand, for Islam (1995) and CEL (1996):

$$\ln A(t) = \ln A(0) + gt \quad (18)$$

Thus, MRW (1992) will estimate equation (19),

$$\begin{aligned} \ln y(t_2) - \ln y(t_1) &= (1 - e^{-\lambda\tau}) \frac{\alpha}{(1 - \alpha - \beta)} \ln s_k \\ &+ (1 - e^{-\lambda\tau}) \frac{\beta}{(1 - \alpha - \beta)} \ln s_h - (1 - e^{-\lambda\tau}) \frac{\alpha + \beta}{(1 - \alpha - \beta)} \ln(n + g + \delta) - \\ &(1 - e^{-\lambda\tau}) \ln y(t_1) + (1 - e^{-\lambda\tau})a + (1 - e^{-\lambda\tau})\epsilon \end{aligned} \quad (19)$$

while Islam (1995) and CEL (1996) will estimate equation (20) below:

$$\begin{aligned} \ln y(t_2) &= (1 - e^{-\lambda\tau}) \frac{\alpha}{(1 - \alpha - \beta)} \ln s_k \\ &+ (1 - e^{-\lambda\tau}) \frac{\beta}{(1 - \alpha - \beta)} \ln s_h - (1 - e^{-\lambda\tau}) \frac{\alpha + \beta}{(1 - \alpha - \beta)} \ln(n + g + \delta) \\ &+ e^{-\lambda\tau} \ln y(t_1) + (1 - e^{-\lambda\tau}) \ln A(0) + (t_2 - e^{-\lambda\tau}t_1)g \end{aligned} \quad (20)$$

Rewriting equation (20) in panel data notation we have basically the same as equation (14) with two main differences: first, the interpretation of the coefficients in the regression changed;

second, we now have three explanatory variables besides the lagged dependent variable, savings on human capital $\ln s_k$ and savings of physical capital variable $\ln s_k$ (instead of only $\ln s$) and $\ln(n + g + \delta)$

2.3 Modified Solow Model à la Barro (1991)

Neoclassical growth models are based on the mechanism of decreasing marginal returns to capital. Poor countries, with lower capital-labor ratios, grow faster than rich countries in these models as the higher marginal productivity of capital induces greater rates of investment. However, faster growth in poor countries seems to be inconsistent with cross-country evidence: in a broad cross section of countries, per capita income growth rates have little or no correlation with initial levels of per capita GDP. Empirical work found evidence of absolute convergence only for smaller and more homogeneous subgroups of countries, such as the OECD.

Two ideas emerged as a solution to the empirical absence of absolute convergence. On the one hand, endogenous growth theory postulated that rates of economic growth depended on other factors, such as the aggregate or average levels of capital and population. These other factor countered the tendency for the marginal product of capital to decrease with accumulation. Convergence was not to be expected.

The other response to lack of absolute convergence was to introduce controls for other determinants of country steady states and look for conditional rather than absolute convergence. Barro (1991) introduces variables such as the school enrollment rate, the fertility rate, the share of government expenditures, political instability, market distortions and dummies for sub-Saharan Africa and Latin America. These factors are meant to proxy for market distortions and, once they are taken into account, Barro finds that poor countries indeed grow faster. This conditional convergence result has proven remarkably strong.

The equation Barro estimates can be represented as

$$\ln y_t - \ln y_{t-1} = a + \beta \ln y_{t-1} + x_t \delta + v_t \quad (21)$$

where $\ln y_t - \ln y_{t-1}$ is the log of the growth rate from 1960 to 1985; a is a constant, $\ln y_t$ is output per capita at t and x_t is a matrix of explanatory variables which include stock and flow variables. The stock variables (e.g. education, capital stock etc.) are measured at the beginning of the sample period (1965) while the flow variables (investment rate etc.) are an average over the period of twenty five years. Specifically, the variables included in x_t are:

i) Measures of human capital: primary-school enrollment rate, secondary school enrollment rate, total fertility rate and child mortality. Both the quality and the quantity of human capital

should be taken into account in growth studies, as they can influence the country steady-state. Countries with high human capital have low fertility rates, low child mortality and high ratios of physical investment to GDP.

ii) Government expenditures: the ratios of real government consumption expenditure to real GDP and public investment to GDP. In previous studies (Barro(1989, 1990)) it was found that the ratio of real government consumption expenditure had a negative relation with growth and investment. The reason is that it lowers savings and growth through distorting effects from taxation.

iii) Political Instability: number of revolutions and coups per year and number per million population of political assassinations per year. Political instability variables measure how well the countries' institutions function. In other words, whether they adversely influence property rights and thus reduce investment and growth.

iv) Economic system. The inclusion of dummies for primarily socialist, mixed between socialist and free enterprise and primarily free enterprise countries intends to capture how economic freedom matters for growth and investment.

v) Market distortions. Distortions of the functioning of markets should impact growth negatively. As a measure of distortions, Barro(1991) considered the purchasing-power-parity (PPP) price index for investment goods. The PPP index is related to the level of economic development.

vi) Dummies for sub-Saharan Africa and Latin America. Due to the poorer growth performance of the countries located in these regions, Barro(1991) includes dummy variables, to account for unobservables that may hinder growth in these regions.

In his later book with Sala-i-Martin (1995-Ch.12), Barro explores the role of other variables in explaining cross-country growth, such as life expectancy and the black-market premium. However, all these other variables are basically different ways of measuring the six important factors listed above that help explain cross-country growth patterns.

3 Estimation and Results

The three models described in the previous section were estimated by different authors using different econometric procedures, summarized in Table 1 below.

Authors/Models	Solow	Aug.Solow	À la Barro.
Kormedi and Maguire (1985)	Cross-Section		
Baumol (1986)	Cross-Section		
De Long (1988)	Cross-Section		
Barro (1991)			Cross-Sec.:GLS
Mankiw,Romer and Weil (1992)	Cross-Section:OLS	Cross-Section:OLS	
Levine and Renelt (1992)			
King and Levine (1993)			
Knight, Loayza, Villaneuva (1993)	Panel:Min.Distance	Panel:Min.Distance	
Loayza (1994)	Panel:Min.Distance	Panel:Min.Distance	
Islam (1995)	Panel:LSDV/Min.Dist.	Panel:LSDV/Min.Dist.	
Caselli,Esquivel,Lefort (1996)	Panel:GMM	Panel:GMM	Panel:GMM

Table 1: Summary of the Empirical Work

3.1 Discussion of the Econometric Procedures

3.1.1 Cross-section studies

The problems of the empirical work using cross-section, such as MRW (1992) and Barro (1991), irrespective of whether the models are estimated by OLS or GLS are:

(i) Small sample size: some estimates are derived from the OECD sample, which includes no more than 22 observations.

(ii) Omission of the individual country-effect:⁶ the partial correlation coefficient between the country-specific effect and the initial value of per capita income is likely to be positive. In other words, the correlation between the individual country-effect and the initial per capita income net of the effect of other variables (savings rate, population growth rate, etc.) is positive. The reason is that the higher individual country-effect tends to be related with higher steady-states. As a consequence, there is an upward bias in the initial income coefficient γ of equation (14).

$$E(\mu_i, y_{i,t1}) > 0 \quad (22)$$

The relation between the coefficient γ and the rate of convergence λ is given by:

⁶Note that, one cannot have individual effects (a dummy for each country) considered in a cross-section.

$$\gamma = e^{-\lambda\tau} \quad (23)$$

or, equivalently,

$$\lambda = \frac{1}{\tau} \ln \gamma \quad (24)$$

Hence, an upward bias in the coefficient of per capita income implies a downward bias in the rate of convergence λ .

(iii) Endogeneity of explanatory variables: as an example, the rate of investment in physical capital is determined simultaneously with the rate of growth in the period. Besides, there is a lagged dependent variable on the right hand side which is correlated with the errors.

3.1.2 Panel data studies

The purpose of panel data is to allow for individual effects to correct for omitted variable bias that would lead to inconsistent estimates. The problems of some of the empirical work using panel data are:

(i) Assumption of a random-effect model. Barro and Sala-i-Martin (1995-ch.12) assume a random effect model when estimating Barro's model. The presence of lagged dependent variable makes the random-effect assumption invalid. The errors will be correlated with right-hand side variables.

Barro and Sala-i-Martin (1995-ch.12) do not explain clearly the econometric procedure they use CEL (1996) and Barro and Sala-i-Martin (1995, ch.12) for example, make some progress on the issue of endogeneity, on the other hand, give the following explanation for the estimation methods used by the authors:

“Barro and Lee (1994a, 1994. They use a panel in which the time series information is derived by splitting the time-period of analysis into two ten-year subperiods—namely, 1965 through 1975 and 1975 through 1985. Stock variables are, respectively, dated 1965 in the equation for 1965-1975 growth and 1975 for 1975-1985 growth. They stack the two cross-sections for the two subperiods and apply a GLS estimator (to correct for serial correlation) where potentially endogenous variables are instrumented by their lagged values. Hence, their work partially corrects the endogeneity problem we found in cross-section work. However, their solution is consistent only under the assumption of random individual effects—individual effects that are correlated over time but not with the other regressors. The problem is that, as showed

above, the presence of lagged dependent variable necessarily makes the random-effect assumption invalid ".(CEL (1996), p. 368).

(ii) Assumption of strict exogeneity of the explanatory variables. Knight, Loayza, Villa-neuva (1993), Loayza (1994) and Islam (1995) employ the procedure proposed by Chamberlain (1984) generally referred to as the π -matrix approach. The assumption required for using this instrumental is the strict exogeneity of the explanatory variables:

$$E(x_{it}v_{is}) = 0, \forall \quad s, t \quad (25)$$

The problem with the use of this estimation method is that the minimum distance parameters will be inconsistent due to the violation of the strict exogeneity assumption. According to CEL (1996) the assumption that is verified is a weaker one, predeterminacy, which is stated as:

$$E(x_{it}v_{is}) = 0, s \geq t \quad (26)$$

$$E(x_{it}v_{is}) \neq 0, s < t \quad (27)$$

In order to obtain consistent estimates of the coefficients for the Solow, Augmented Solow and the modified Solow model, CEL (1996) assume that the following assumptions are valid:

(A1) No first order serial correlation of the residuals:

$$E(v_{it}v_{it-1}) = 0 \quad (28)$$

(A2) Stock variables in the matrix of explanatory variables are predetermined.⁷

(A3) Flow variables in the matrix of explanatory variables are not predetermined for the current v_{it} but are predetermined for the leads v_{it+1}, v_{it+2} , etc..

Based on assumptions (A1) to (A3) the authors estimate the Solow growth model (equation (14) is reproduced below) in the following way ⁸:

$$y_{it} = \gamma y_{i,t-1} + \sum_{j=1}^2 \beta_j x_{it}^j + \eta_t + \mu_i + v_{it} \quad (29)$$

a) eliminating the time component η_t by expressing the variables y_{it} and x_{it}^j in deviations from period means. As an example, for y_{it} we have:

⁷These are the variables measured at the beginning of the period.

⁸We will use equation (14) to illustrate the steps CEL(1996) followed in their estimation of the Solow model, Augmented Solow and the modified Solow.

$$\tilde{y}_{it} = y_{it} - \bar{y}_t \quad (30)$$

b) eliminating the country-individual effect by taking first differences;

$$\Delta \tilde{y}_{it} = \gamma \Delta \tilde{y}_{it-1} + \beta_0 \Delta \tilde{x}_{it}^1 + \beta_1 \Delta \tilde{x}_{it}^2 + \Delta \tilde{v}_{it} \quad (31)$$

c) using instrumental variables to correct for endogeneity in the case of flow variables in \tilde{x}_{it}^j . The instruments used are all past values of the explanatory variable. The stock variables in \tilde{x}_{it}^j are measured at $t - 1$ (beginning of period) and thus they do not have to be instrumented.

d) Generalized Method of Moments estimation is used in order to get consistent and efficient estimators.

Let's rewrite equation (31) in a different manner:

$$\Delta y_{it} = \gamma \Delta y_{it-1} + \beta_0 \Delta x_{it-1}^s + \beta_1 \Delta x_{it}^f + \Delta v_{it} \quad (32)$$

$$x_{it}^f \equiv \text{avg}(\text{flow}_{it}, \text{flow}_{it-1}) \quad (33)$$

Assume that x_{it-1}^s are the stock variables which are measured at $t - 1$ and that x_{it}^f are the flow variables which are measured as an average from $t - 1$ to t . Given (A1)-(A3) we have that:

$$E^*(v_{it}/y_{it-1}, x_{it-1}^s, x_{it-2}^f) = 0 \quad (34)$$

Equation (34) implies that:

$$E(y_{it-s} \Delta v_{it}) = 0, s \geq 2 \quad (35)$$

$$E(x_{it-s}^s \Delta v_{it}) = 0, s \geq 2 \quad (36)$$

$$E(x_{it-s}^f \Delta v_{it}) = 0, s \geq 3 \quad (37)$$

where in equation (36):

$$x_{it-3}^f = \text{avg}(\text{flow}_{it-2}, \text{flow}_{it-3}) \quad (38)$$

Based on (35), (36) and (37) we construct the instrument matrix Z_i . Given that $t = 0, 1, 2, 3, 4, 5$ (corresponding to the years 1960, 1965, 1970, 1975, 1980 and 1985) we lose the first two observations as we wrote the system in first differences and we are instrumenting the variables with their lagged values.

$$t = 2 \implies z_i^1 = [y_{i0}, x_{i0}^s] \quad (39)$$

$$t = 3 \implies z_i^2 = [y_{i0}, x_{i0}^s, x_{i0}^f, y_{i1}, x_{i1}^s] \quad (40)$$

$$t = 4 \implies z_i^3 = [y_{i0}, x_{i0}^s, x_{i0}^f, y_{i1}, x_{i1}^s, x_{i1}^f, y_{i2}, x_{i2}^s] \quad (41)$$

$$t = 5 \implies z_i^4 = [y_{i0}, x_{i0}^s, x_{i0}^f, y_{i1}, x_{i1}^s, x_{i1}^f, y_{i2}, x_{i2}^s, x_{i2}^f, y_{i3}, x_{i3}^s] \quad (42)$$

$$Z_i = \text{diag}(z_i^1 z_i^2 z_i^3 z_i^4) \quad (43)$$

Let's call:

$$V_i = [v_{i1} - v_{i0}, v_{i2} - v_{i1}, v_{i3} - v_{i2}, v_{i4} - v_{i3}] \quad (44)$$

Thus, we have that

$$E(Z_i' V_i) = 0 \quad (45)$$

CEL (1996) apply GMM to obtain consistent as well as efficient estimators. In the Appendix we explain the method used by the authors. After obtaining the estimates, CEL (1996) conduct several tests of specification, which we describe in the Appendix.

3.2 Results

In Table 2 we compare the rates of convergence and the share of capital to output estimations obtained in MRW (1992), Islam (1995) and CEL (1996) and in Table 3 we compare the rates of convergence and the share of capital to output estimations obtained by the same authors for the Augmented Solow model. Finally, in Table 4 we compare the results obtained from estimations of the modified model à-la-Barro.

Basically, the main message of this section is that the results vary a lot from different econometric procedures. The results of some tests of specification for the Solow and the Augmented Solow are reported below. The Sargan test does not reject the validity of the overidentifying restrictions. The Difference Sargan test as well as the m_2 test do not reject the null hypothesis of no second order residual correlation as seen in Tables 5 and 6.⁹

4 Conclusion

This paper reviews the theoretical growth models as well as developments in the econometric methodology on panel data estimation as applied to the study of growth empirics. We show that

⁹See the Appendix for a detailed explanation of the tests.

the Generalized Method of Moments (GMM) allows us to obtain consistent as well as efficient estimators of the coefficients in panel data empirical growth models. The main suggestion is the potential benefit of studying regional income convergence rather than convergence across countries. There are several factors shared among regions of the same country- including, but not limited to factor mobility and institutions - that affect the rate of convergence in predictable ways. The application of the econometric instrumental presented above may bring new insights to the analysis of growth empirics.

References

- [1] Arellano, M. and S. Bond (1991), “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations”, *Review of Economic Studies* 58, 277-297.
- [2] Barro, R.J. (1991), “Economic Growth in a Cross-section of Countries”, *Quarterly Journal of Economics* 106, 407-443.
- [3] Barro, R.J. and X. Sala-i-Martin (1995), *Economic Growth*, New York: McGraw-Hill.
- [4] Caselli, F., G. Esquivel and F. Lefort (1996), “Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics”, *Journal of Economic Growth* 1:363-389.
- [5] Chamberlain, G. (1984), “Panel Data ”, in Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, v.2, (pp.1247-1313).
- [6] Hirano, K. (1999), Lecture notes. Mimeo, University of California Los Angeles.
- [7] Islam, N. (1995), “Growth Empirics: A Panel Data Approach”, *Quarterly Journal of Economics* 110, 1127-1170.
- [8] Knight, M. , N. Loayza, and D. Villaneuva (1993), “Testing the Neoclassical Growth Model ”. *IMF Staff Papers*, 40, 512-541.
- [9] Loayza, N. (1994), “A Test of the International Convergence Hypothesis Using Panel Data”. *Policy Research Working Paper* No.1333, The World Bank.
- [10] Mankiw, N.G.D. Romer and D. Weil (1992), “A Contribution to the Empirics of Economic Growth” , *Quarterly Journal of Economics* 107, 407-437.
- [11] Romer, D. (1996), “Advanced Macroeconomics”. New York: McGraw-Hill.
- [12] Solow, R. (1956), “A Contribution to the Theory of Economic Growth ”. *Quarterly Journal of Economics* 70, 65-94.

A Appendix

A.1 Solow Model

The production function is given by:

$$Y(t) = K(t)^\alpha [A(t)L(t)]^{1-\alpha}, 0 < \alpha < 1 \quad (46)$$

where t is time, $Y(t)$ output, $K(t)$ capital stock, $L(t)$ labor, $A(t)$ technological progress and α the share of capital on output.

Assume that the population growth rate is given by n and the rate of technological progress by g . Thus,

$$L(t) = L(0)e^{nt} \quad (47)$$

$$A(t) = A(0)e^{gt} \quad (48)$$

Let's define:

$$\hat{k}(t) = \frac{K(t)}{A(t)L(t)} \quad (49)$$

$$\hat{y}(t) = \frac{Y(t)}{A(t)L(t)} \quad (50)$$

Thus,

$$\hat{y}(t) = \hat{k}(t)^\alpha \quad (51)$$

Differentiating $\hat{k}(t)$ with respect to time:¹⁰

$$\dot{\hat{k}}(t) = \frac{\dot{K}(t)}{A(t)L(t)} - \frac{K(t)}{A(t)L(t)} \left(\frac{\dot{A(t)L(t)}}{A(t)L(t)} \right) \quad (52)$$

Given that:

$$K_{t+1} = (1 - \delta)K_t + I_t \quad (53)$$

where I_t is investment. We have that $K_{t+1} - K_t = -\delta K_t + I_t$ in discrete time which is equivalent to $\dot{K}(t) = -\delta K_t + I_t$ in continuous time. Dividing by $A(t)L(t)$ and given that $I_t = sY(t)$ we have:

¹⁰Note that:

$$\hat{k}(t) = \frac{K(t)}{A(t)L(t)}$$

In order to totally differentiate $\hat{k}(t)$ we should apply the rule $\frac{u}{v} = \frac{u'v - uv'}{v^2}$ which is equivalent to $\frac{u}{v} = \frac{u'}{v} - \frac{u}{v} \frac{v'}{v}$ where $\frac{u}{v} = \frac{K(t)}{A(t)L(t)}$.

$$\frac{\dot{K}(t)}{A(t)L(t)} = -\frac{\delta K(t)}{A(t)L(t)} + \frac{sY(t)}{A(t)L(t)} \quad (54)$$

which can be re-written as: $\frac{\dot{K}(t)}{A(t)L(t)} = -\delta \hat{k}(t) + s\hat{y}(t)$. Replacing in equation (52) we get:

$$\dot{\hat{k}}(t) = -\delta \hat{k}(t) + s\hat{y}(t) - \hat{k}(t) \left(\frac{\dot{A(t)L(t)}}{A(t)L(t)} \right) \quad (55)$$

Given that:

$$\frac{\dot{A(t)L(t)}}{A(t)L(t)} = \frac{\dot{A(t)L(t)}}{A(t)L(t)} + \frac{\dot{A(t)L(t)}}{A(t)L(t)} \quad (56)$$

$$\frac{\dot{A(t)L(t)}}{A(t)L(t)} = g + n \quad (57)$$

Replacing equation (57) in equation (55) we get:

$$\dot{\hat{k}}(t) = s\hat{y}(t) - (n + g + \delta)\hat{k}(t) \quad (58)$$

$$\dot{\hat{k}}(t) = s\hat{k}(t)^\alpha - (n + g + \delta)\hat{k}(t) \quad (59)$$

where $s = \frac{\dot{K}(t) + \delta K(t)}{Y(t)}$.

In the steady-state, $\dot{\hat{k}}(t) = 0$.

Thus, from (51) and (59) we obtain the steady-state values of \hat{k}^* and \hat{y}^* , respectively:

$$\hat{k}^* = \left(\frac{s}{n + g + \delta} \right)^{\frac{1}{1-\alpha}} \quad (60)$$

$$\hat{y}^* = \left(\frac{s}{n + g + \delta} \right)^{\frac{\alpha}{1-\alpha}} \quad (61)$$

Note that equation (60) was obtained by assuming $\dot{\hat{k}}(t) = 0$ in equation (59):

$$s = (n + g + \delta) \hat{k}^{*1-\alpha}$$

We can get the speed of convergence by linearizing around the steady-state \hat{k}^* (or \hat{y}^*). A first-order-Taylor-series approximation of $\dot{\hat{k}}(t)$ around $\hat{k}(t) = \hat{k}^*$ is:

$$\dot{\hat{k}}(t) \simeq \left(\frac{\partial \dot{\hat{k}}(t)}{\partial \hat{k}(t)} \Big|_{\hat{k}(t)=\hat{k}^*} \right) (\hat{k}(t) - \hat{k}^*) \quad (62)$$

Equation (62) says that $\dot{\hat{k}}(t)$ is approximately equal to the difference between $\hat{k}(t)$ and \hat{k}^* times the derivative of $\dot{\hat{k}}(t)$ with respect to $\hat{k}(t)$ evaluated at $\hat{k}(t) = \hat{k}^*$. First, we proceed by calculating the part inside the parenthesis in equation (62):

$$\left. \frac{\partial \dot{\hat{k}}(t)}{\partial \hat{k}(t)} \right|_{\hat{k}(t)=\hat{k}^*} = s\alpha \hat{k}(t)^{\alpha-1} - (n + g + \delta) \quad (63)$$

Evaluating the equation above at $\hat{k}(t) = \hat{k}^*$ and by using equation (??) we obtain:

$$\left. \frac{\partial \dot{\hat{k}}(t)}{\partial \hat{k}(t)} \right|_{\hat{k}(t)=\hat{k}^*} = -(1 - \alpha)(n + g + \delta) \quad (64)$$

Second, by substituting (64) in (62) we obtain:

$$\dot{\hat{k}}(t) \simeq -(1 - \alpha)(n + g + \delta)(\hat{k}(t) - \hat{k}^*) \quad (65)$$

Finally, by defining $x(t) = \hat{k}(t) - \hat{k}^*$ and $\lambda = (1 - \alpha)(n + g + \delta)$ we have that equation (65) is equivalent to $\dot{x}(t) \simeq -\lambda x(t)$. As the growth rate of x is constant and equal to $-\lambda$, we obtain:

$$x(t) \simeq x(0)e^{-\lambda t} \quad (66)$$

which can be written in terms of \hat{k} :

$$\hat{k}(t) - \hat{k}^* \simeq e^{-\lambda t}(\hat{k}(0) - \hat{k}^*) \quad (67)$$

where $\lambda = (1 - \alpha)(n + g + \delta)$.¹¹

By the same procedure, it can be shown that:

$$\hat{y}(t) - \hat{y}^* \simeq e^{-\lambda t}(\hat{y}(0) - \hat{y}^*) \quad (68)$$

where $\lambda = (1 - \alpha)(n + g + \delta)$.

We can manipulate equation (68) to obtain:

$$\ln \hat{y}(t_2) = (1 - e^{-\lambda \tau}) \ln \hat{y}^* + e^{-\lambda \tau} \ln \hat{y}(t_1) \quad (69)$$

where $\tau = t_2 - t_1$ Subtracting $\ln \hat{y}(t_1)$ from both sides:

$$\ln \hat{y}(t_2) - \ln \hat{y}(t_1) = (1 - e^{-\lambda \tau}) \ln \hat{y}^* - (1 - e^{-\lambda \tau}) \ln \hat{y}(t_1) \quad (70)$$

¹¹This explanation is based on Romer (1996).

Replacing the steady-state \hat{y}^* given by (57) in equation (70) we get:

$$\begin{aligned} \ln \hat{y}(t_2) - \ln \hat{y}(t_1) &= (1 - e^{-\lambda\tau}) \frac{\alpha}{(1 - \alpha)} \ln s - \\ &(1 - e^{-\lambda\tau}) \frac{\alpha}{(1 - \alpha)} \left[\ln(n + g + \delta) - (1 - e^{-\lambda\tau}) \ln \hat{y}(t_1) \right] \end{aligned} \quad (71)$$

A.2 Methods of Estimation

A.2.1 Least-Squares

Assumption:

$$E(\varepsilon_t) = 0; V(\varepsilon_t) = \sigma^2, \quad i.i.d. \quad (72)$$

Identifying restriction (the orthogonality condition):

$$E(x_i(y_i - x'_i\delta)) = 0 \quad (73)$$

By the analogy principle:

$$\frac{1}{n} \sum_{i=1}^n x_i(y_i - x'_i\delta) = 0 \quad (74)$$

$$\hat{\delta}_{ols} = (X'X)^{-1}X'y \quad (75)$$

A.2.2 Minimum Distance Estimation

Chamberlain (1982) proposes this estimation procedure.¹²

Assumption (strict exogeneity of the explanatory variables):

$$E(x_{it}, v_{is}) = 0, \forall \quad s, t \quad (76)$$

The Solow model is (assume the restricted model where there is only one explanatory variable x_{it} besides the lagged dependent variable y_{it-1} (equation (14)):

$$y_{it} = \gamma y_{i,t-1} + \beta x_{it} + \mu_i + v_{it} \quad (77)$$

where $x_{it} = [\ln(s) - \ln(n + g + \delta)]$.

¹²This explanation is based on Islam (1995-p.1167).

Let's ignore the time effect η_t by using time dummies.

Assume the following specification for μ_i and y_{i0} :

$$\mu_i = k_0 + k_1x_{i1} + k_2x_{i2} + \dots + k_Tx_{iT} + \Psi_i \quad (78)$$

$$y_{i0} = \Phi_0 + \Phi_1x_{i1} + \Phi_2x_{i2} + \dots + \Phi_Tx_{iT} + \xi_i \quad (79)$$

where $E(\Psi_i | x_{i1}, \dots, x_{iT}) = 0$; $E(\xi_i | x_{i1}, \dots, x_{iT}) = 0$.

The MD estimator is based on substituting out the lagged dependent variable.

As an example, assume $T = 3$ (years 1965, 1970, 1975). We will have (ignoring the subscripts i):

$$y_1 = \gamma y_0 + \beta x_1 + \mu + v_1 \quad (80)$$

$$y_2 = \gamma^2 y_0 + \gamma\beta x_1 + \beta x_2 + (\mu + \gamma\mu) + (v_2 + \gamma v_1) \quad (81)$$

$$y_3 = \gamma^3 y_0 + \gamma^3\beta x_1 + \gamma^2\beta x_2 + \beta x_3 + (\mu + \gamma\mu + \gamma^2\mu) \quad (82)$$

$$+ (v_3 + \gamma v_2 + \gamma^2 v_1) \quad (83)$$

In matrix form we will have:

$$\begin{aligned} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} &= \begin{bmatrix} \gamma \\ \gamma^2 \\ \gamma^3 \end{bmatrix} y_0 + \begin{bmatrix} \beta & 0 & 0 \\ \gamma\beta & \beta & 0 \\ \gamma^3\beta & \gamma^2\beta & \beta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &+ \begin{bmatrix} 1 \\ 1 + \gamma \\ 1 + \gamma + \gamma^2 \end{bmatrix} \mu + \begin{bmatrix} v_1 \\ v_2 + \gamma v_1 \\ v_3 + \gamma v_2 + \gamma^2 v_1 \end{bmatrix} \end{aligned} \quad (84)$$

If we substitute μ_i and y_{i0} according to equations (78) and (79) in the matrix above, we will get the π -matrix of the reduced form coefficients as:

$$\pi = \begin{bmatrix} \gamma \\ \gamma^2 \\ \gamma^3 \end{bmatrix} \Phi' + \begin{bmatrix} \beta & 0 & 0 \\ \gamma\beta & \beta & 0 \\ \gamma^3\beta & \gamma^2\beta & \beta \end{bmatrix} + \begin{bmatrix} 1 \\ 1 + \gamma \\ 1 + \gamma + \gamma^2 \end{bmatrix} K' \quad (85)$$

where

$$\begin{aligned} \Phi' &= \begin{bmatrix} \Phi_1 & \Phi_2 & \Phi_3 \end{bmatrix} \\ K &= \begin{bmatrix} k_1 & k_2 & k_3 \end{bmatrix} \end{aligned}$$

and the π -matrix is the matrix of coefficients of x_1 , x_2 and x_3 .

If the x 's are strictly exogenous, then they are uncorrelated with the errors and you can estimate the reduced form coefficients by least-squares (OLS).

Note that the π -matrix above is 3×3 (it has nine elements to be estimated) which are non-linear functions of the eight coefficients $(\gamma, \beta, \Phi_1, \Phi_2, \Phi_3, k_1, k_2, k_3)$.¹³

In this case, the best way to estimate the eight coefficients by using nine non-linear functions is to use the minimum distance. If we call $\hat{\theta}$ the vector of the eight coefficients we want to estimate we have:

$$\hat{\theta} = \arg \min (vec \pi - g(\theta))' A_n^{-1} (vec \pi - g(\theta)) \quad (86)$$

where $g(\theta)$ maps the elements of θ into $vec \pi$. The optimal choice of A_n^{-1} is the inverse of Ω which by the analogy principle can be estimated by its consistent sample analog $\hat{\Omega}$. Note that Ω is a heteroskedasticity-consistent weighing matrix:

$$\Omega = E[(y_i - \pi^0 x_i)(y_i - \pi^0 x_i)' \otimes \phi_x^{-1}(x_i x_i') \phi_x^{-1}] \quad (87)$$

where Ω is the true coefficients matrix and $\phi_x = E(x_i x_i')$. On the other hand, $\hat{\Omega}$ is:

$$\hat{\Omega} = \frac{1}{N} \sum [(y_i - \pi x_i)(y_i - \pi x_i)' \otimes S_x^{-1}(x_i x_i') S_x^{-1}] \quad (88)$$

where $S_x = \frac{\sum_i^n x_i x_i'}{N}$.

You could use the DFP algorithm to solve the minimization problem above.

A.2.3 Generalized Method of Moments Estimation (GMM)

Assumptions: (A1)-(A3) in section 3.3.1.

Identifying restriction (the orthogonality condition):

$$E(Z_i' V_i) = 0 \quad (89)$$

By the analogy principle:

$$\frac{1}{N} \sum_{i=1}^n Z_i' V_i = 0 \quad (90)$$

¹³As Islam(1995-p.1169) points out, in the case of T=5 the Π - matrix has 25 elements and the number of coefficients to be estimated are 12.

The GMM estimator of β_{gmm}^j is:

$$\beta_{gmm}^j = \left(\frac{1}{n} \sum_{i=1}^n Z_i' V_i \right)' A_j \left(\frac{1}{n} \sum_{i=1}^n Z_i' V_i \right) \quad (91)$$

which can also be expressed as:

$$\beta_{gmm}^j = (X' Z A_j Z' X)^{-1} X' Z A_j Z' Y \quad (92)$$

where

$$\begin{aligned} Z &= [Z_1', Z_2', \dots, Z_n']'_{nT \times m} \\ X &= [\Delta X_1', \Delta X_2', \dots, \Delta X_n']'_{nT \times k}, \quad X_i = [X_{i,0}', \dots, X_{i,T\tau}'] \\ Y &= [\Delta Y_1', \Delta Y_2', \dots, \Delta Y_n']'_{nT \times 1}, \quad Y_i = [y_{i0}, \dots, y_{i,T\tau}]' \end{aligned}$$

Note that the estimated residuals are:

$$\hat{V}_i \equiv \Delta Y_i - \Delta X_i \beta_{gmm}^j \quad (93)$$

The two-step procedure was used to compute the GMM estimators. In the first step, it was assumed that V_i are independent and identically distributed with constant variance σ_v^2 . This implies that

$$E(\hat{V}_i \hat{V}_i') = \sigma_v^2 H \quad (94)$$

where H is a matrix with 2 in the principal diagonal, -1 in the secondary diagonal and 0 otherwise.

The matrix H can be understood if we observe that assumption (A1) in section 3.3.1 implies:

$$E(\Delta \hat{v}_{it})^2 = E(\hat{v}_{it}^2 - 2\hat{v}_{it}\hat{v}_{it-1} + \hat{v}_{it-1}^2) = 2\sigma_v^2 \quad (95)$$

$$E(\Delta \hat{v}_{it} \Delta \hat{v}_{it-1}) = E[(\hat{v}_{it} - \hat{v}_{it-1})(\hat{v}_{it-1} - \hat{v}_{it-2})] = -\sigma_v^2 \quad (96)$$

In the first stage we have that

$$A_1 = \frac{1}{n} \sum_{i=1}^n (Z_i' H Z_i)^{-1} \quad (97)$$

In the second stage, homoscedasticity is relaxed. The β_{gmm}^j are recalculated based on $\Delta \hat{v}_{it}$ obtained in the first stage using equation (93) and using the weight matrix A_2 constructed as:

$$A_2 = \frac{1}{n} \sum_{i=1}^n (Z_i' \hat{V}_i^{-1} \hat{V}_i^{-1'} Z_i)^{-1} \quad (98)$$

A.3 Tests of Specification

A.3.1 Sargan Test or Test of Overidentifying Restrictions

The Sargan test evaluates the overall validity of the identifying restrictions. In other words, it tests if the instruments used in the estimation process are valid ones:

$$H_0 : E(Z_i' V_i) = 0$$

Given that:

$$V_i = [v_{i1} - v_{i0}, v_{i2} - v_{i1}, v_{i3} - v_{i2}, v_{i4} - v_{i3}] \quad (99)$$

and

$$\hat{V} = [\hat{V}_1' \dots \hat{V}_n'] \quad (100)$$

The Sargan test statistic is given by:

$$s = \hat{V}' Z \left(\sum_{i=1}^n Z_i' \hat{V}_i \hat{V}_i' Z_i \right)^{-1} Z' \hat{V} \quad (101)$$

The Sargan test statistic asymptotic distribution is χ_{M-K}^2 where M is the number of columns in X and K is the number of columns in Z . These matrices are described in section 3.3.1 of the paper.

When we reject the null hypothesis in a Sargan test we do not know which are the identifying restrictions that are correct.

A.3.2 Difference Sargan

CEL (1996) assumed that:

(A1) no first order serial correlation of the residuals¹⁴:

$$E(v_{it}, v_{it-1}) = 0 \quad (102)$$

The null hypothesis of the difference Sargan test is that there is no serial correlation in the errors in levels. The test statistic is:

¹⁴The authors did not explain how they constructed the difference Sargan test.

$$d_s = s - s_I \overset{A}{\sim} \chi_{p-p_1}^2 \quad (103)$$

where:

- s is the Sargan test statistic calculated assuming that the errors are not second order serially correlated. This means that y_{it-2} are valid instruments (as well as y_{it-3} , y_{it-4} and so on), i.e.

$$E(y_{it-j}\Delta v_{it}) = 0, j \geq 0 \quad (104)$$

- s_I is the Sargan test statistic calculated limiting the number of instruments to the ones that are not second order serially correlated. This means that y_{it-3} are valid instruments (as well as y_{it-4} , y_{it-5} and so on).

A.3.3 m_2 Test

The m_2 test evaluates if there is no second order residual autocorrelation

$$E(\Delta v_{it}\Delta v_{it-2}) \neq 0 \quad (105)$$

The null hypothesis of the m_2 test is:

$$H_0 : E(v_{it} - v_{it-1}, v_{it-2} - v_{it-3}) = 0 \quad (106)$$

which is equivalent to

$$H_0 : E(v_{it}v_{it-2}) - E(v_{it}v_{it-3}) - E(v_{it-1}v_{it-2}) + E(v_{it-1}v_{it-3}) = 0 \quad (107)$$

The m_2 test statistic is:

$$m_2 = \frac{\hat{V}_{-2}' \hat{V}_*}{Q} \overset{A}{\sim} N(0, 1) \quad (108)$$

where Q is the appropriate standardization.

$$\begin{aligned} \hat{V}_{-2i} &\equiv [\hat{v}_{i1}, \dots, \hat{v}_{it-2}]' \\ \hat{V}_{-2} &\equiv [\hat{V}'_{-21}, \dots, \hat{V}'_{-2n}]' \\ \hat{V}_{*i} &\equiv [\hat{v}_{i3}, \dots, \hat{v}_{it}]' \\ \hat{V}_* &\equiv [\hat{V}'_{*1}, \dots, \hat{V}'_{*n}]' \end{aligned}$$

The m_2 follows a standard normal distribution¹⁵.

Example:

Assume $t = 3$ and $i = 2$. We will have:

$$\begin{aligned}\hat{V}_{-21} &\equiv [\hat{v}_{11}] \\ \hat{V}_{-22} &\equiv [\hat{v}_{21}] \\ \hat{V}_{-2} &\equiv [\hat{v}_{11} \hat{v}_{21}]' \\ \hat{V}_{*1} &\equiv [\hat{v}_{13}] \\ \hat{V}_{*2} &\equiv [\hat{v}_{23}] \\ \hat{V}_* &\equiv [\hat{v}_{13} \hat{v}_{23}]\end{aligned}$$

Thus,

$$\hat{V}_{-2}' \hat{V}_* = \begin{bmatrix} \hat{v}_{11} & \hat{v}_{21} \end{bmatrix} \begin{bmatrix} \hat{v}_{13} \\ \hat{v}_{23} \end{bmatrix} = \hat{v}_{11} \hat{v}_{13} + \hat{v}_{21} \hat{v}_{23} \quad (109)$$

$$\hat{V}_{-2}' \hat{V}_* = \sum_{t=2}^t \sum_{i=1}^n \hat{v}_{it-2} \hat{v}_{it} \quad (110)$$

i.e, it is an average of second order covariances of residuals.

Note that CEL (1996) assumed no first order serial correlation in the residuals in order to use y_{it-2} as an instrumental.

$$E(y_{it-2} \Delta v_{it}) = 0 \quad (111)$$

If you do not reject the null you still do not know if the errors are not at all serially correlated or if they follow a random walk. To check which one is the case it is necessary to do an m_1 test¹⁶

¹⁵For more details see Arellano and Bond (1991).

¹⁶See Arellano and Bond (1991) for an explanation of the m_1 test.